# 1. Introduction

## 1.1. Practical Problem

In the field of school transportation, the occurrence of bus breakdowns and delays is one big issue, as both of the two situations disrupt the daily routines and schedules of students. This problem becomes especially pronounced when considering the safety of children on the school bus. In recent times, as there are more crimes and accidents in the boarding and operation of school bus[1], parents are increasingly anxious about their children's well-being during their daily commute to school. These concerns range from children falling asleep on the bus to the fear of kids either missing the bus or disembarking at the wrong location. Prolonged delays in pick-up and drop-off times, extended waiting periods for boarding and receiving the students, issues related to tracking and communication, unexplained absences, unusual behavior among children, and cases of students accidentally boarding the wrong bus have all contributed to the worries of parents. Schools, who should be responsible for the students' safety and teaching quality relating to the schedule, have similar concerns because as most of them hire outsourcing employment, they do not have a direct control of the bus situations. According to the above, it is important to figure out what causes the delays and breakdowns, so that we can provide insights to improve the school bus service quality.

There are many reasons for bus breakdowns or delays. For example, when a bus experiences a breakdown during its scheduled trip, it necessitates the rescheduling of one or more vehicles to cover not only that affected trip but also other scheduled routes.[2] Or, the electrical breakdown time delay measurements have been carried out as a function of several parameters, which dominantly influence electrical breakdown time delay.[3] Also, traditional and basic reasons like a traffic jam or a flat tire should be first in the spotlight. We need to know if the general bus situation could be applied in school bus situation.

## 1.2. Research Problem

We have a dataset of school bus breakdowns/delays in New York City. The Bus Breakdown and Delay system collects information from school bus vendors operating out in the field in real-time. When bus staff encounters delays on their routes, they communicate this information to the dispatcher at the bus vendor's central office through radio communication. Subsequently, the bus vendor staff logs into the Bus Breakdown and Delay system.[4] Within this system, they record the specific event, detailing the nature of the breakdown or delay, and this information provides a comprehensive overview of the disruptions faced by the school bus service.

Research is needed to anticipate school bus breakdowns and delays to facilitate proactive measures, identify patterns within operational data to uncover insights into factors contributing to disruptions, and optimize bus routes, timings, and resource allocation to enhance the overall efficiency and ensure students' safety.

## 1.3. Research Objectives

The following objectives are investigated within this research report:

- Which factors contribute to the frequency of delays/breakdowns?
- Is there a hourly fluctuation or seansonality effect for the frequency?

- What about the main patterns in a specific time?
- Which factors contribute to the duration of delay time?
- Which bus companies are responsible for a longer delay time?
- In which area do bus services more likely to experience longer delays?

## 2. Literature Review

The research related to improving the school bus service dates from the last century. School bus routing problem (SBRP) has been studied by Newton et. al.[5] and the later scholars since 1969. Burke et. al. in 1996 study evaluation of the effectiveness of the route in promoting safe behavior among school children boarding school buses.[6] Transportation is an area where operations have a great impact on systems by improving the service quality and reducing operating costs. As for improvement of the school bus system, in 1997 Jeffrey et.al. Investigate various issues related to the development of a computerized system to help route and schedule school buses throughout the five boros of New York City.[7] In 2010 Park et. al. aim to provide a comprehensive review of the school bus routing problem.[8] SBRP seeks to plan an efficient schedule for a fleet of school buses where each bus picks up students from various bus stops and delivers them to their designated schools while satisfying various constraints such as the maximum capacity of a bus, the maximum riding time of a student in a bus, and the time window of a school.

Modern systems or techniques for a real-time school bus service improvement have been developed over time. It is very difficult for us to query and choose the right bus routes. Aiming at this problem in 2014 Shiyao et. al. discuss a kind of public transport inquiry system.[9] This system takes the school bus as the implementation model, and in allusion to the intelligent bus query system based on ZigBee network, GPS and internet to discuss. Notification system is also an advance. The FSR sensors could detect the left child on the school bus accurately and the alarm system can warn timely. In 2019, Liu et. al. study alarm system design of young children being left on school bus based on pressure sensor array.[10] The detection method and the alarm system can be used for young children school bus to improving the safety of children riding. In allusion to the delay issue, technical improvements are needed for such as real-time bus information, controlling run time and headway delay. In 2017 Yue et. al. aim to carry out a preliminary survey to determine the problems of school shuttle bus that faced by the students in a selected educational institution,[11] their perceptions of using shuttle bus tracking and information mobile application and impacts of real-time information of public transits on bus ridership and towards smart mobility solutions.

## 3. Methodology

Our main data mining research methodology is CRISP-DM, which is an industry-independent process model for applying data mining projects.[12] It consists of six iterative phases from business understanding to deployment (see Fig 1). We follow the procedure to conduct our project. First, we collect background information to understand our business and then find relevant databases to explore the data. We may raise some hypothesis and visualize the data to prove it or reject it. Then the initial viewpoint of the data provides us a comprehensive understanding of the context and purpose of the analysis about what we can do on the business. After several iterations, we will have a specific objective. We move on to the data preparation where we clean, transform and organize raw data into a format suitable for modeling. We will apply several models on the data to figure out some patterns inside our data. We may deal with the raw data multiple times for different model selections. Also, the same model with different data preprocessing methods will give a different result. For example, we can either remove the null data, which will lose some information in other data fields, or assign it with the mode value, which gives some factor a higher weight. When we build a model using the most appropriate data mining algorithm by evaluating the models in scores like the accuracy rate, root mean square error etc., we use the model for data mining. All the results of different models contribute to the final result to support our data mining purpose. Several iterations should be considered in each step until we get a robust, conceivable and almost accurate data analysis result. Finally we extract practical insights enabling informed decision-making and strategic planning from the results in the deployment phase. This process improves the business understanding, along with the development of this buisness over time, so we can have one more data mining iteration until the problem is completely solved.

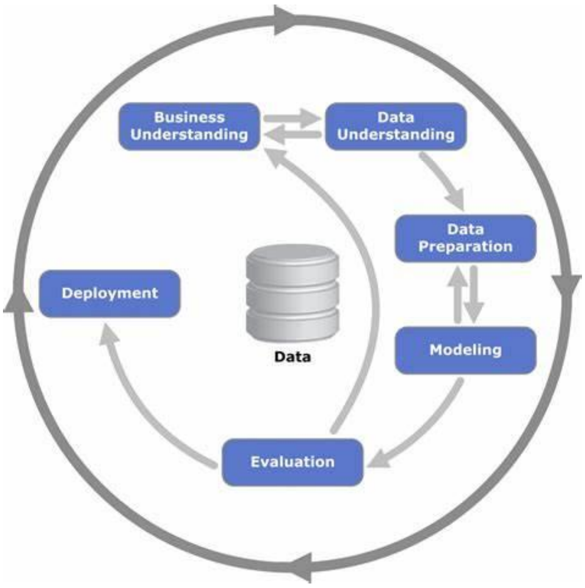


Figure 1. CRISP DM Process

# 4. Modelling

### 4.1. **Design Process**

We use IBM SPSS Modeler, Python with scikit-learn package, and Python with Spark engine to convert data into insights. We utilize IBM SPSS Modeler in our first iteration, as SPSS is known for its user-friendly interface and provides a quick, intuitive approach for data mining. We use Python in our second iteration, as the scikit-learn package offers us the flexibility to apply a wide range of customized data preprocessing and machine learning algorithms. We hope to improve the efficiency of dealing with millions of data records, so Python with Spark comes into play.

To begin, the choice of data sources is a critical consideration. The Bus Breakdown and Delay system has provided useful datasets and claimed that there are manual quality issues inside them. Along with reading the official document, we define the intital research question based on our dataset. The dataset has 21 fields and 360,325 records. The data fields can roughly divided into four type: categorical type including flag type, numeric type, timestamp type and ID type indicating the unique record key. We explore the dataset using several visualisations (part can be seen in Fig 2) and refine our objective questions: a) whether the time related data values, with the other suspected confounders, have a strong correlation with the frequency of the school bus breakdowns or delays, and b) which factors contribute to the duration of delay time. Our following process is tailored to extract the most relevant and valuable information.

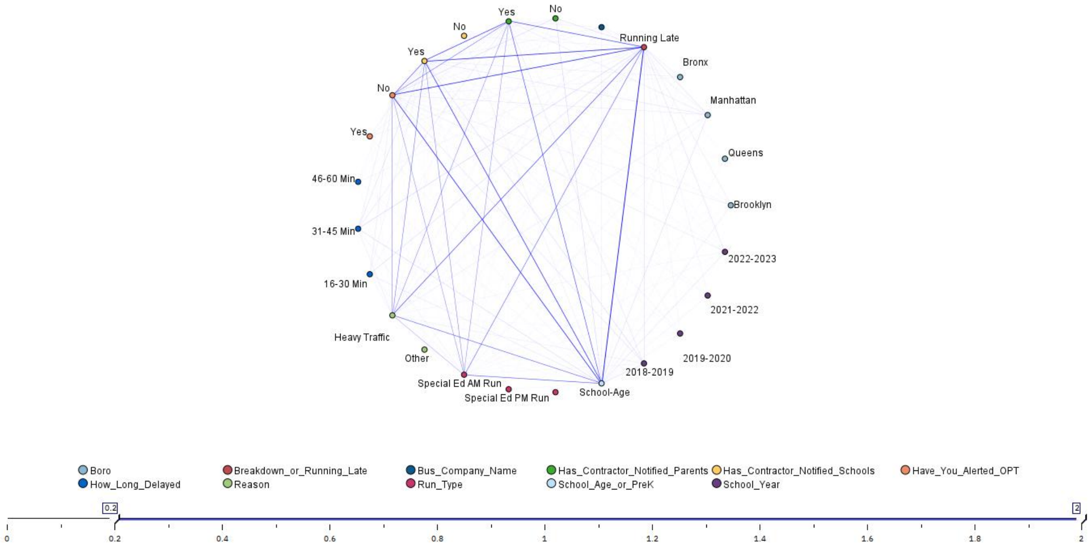


Figure 2a. Data Exploration of Correlations among Factors Based on Frequency

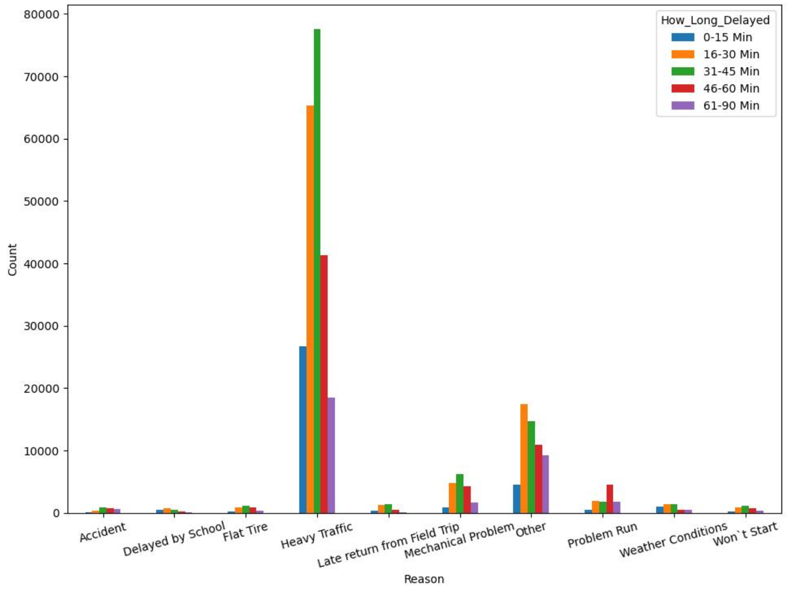
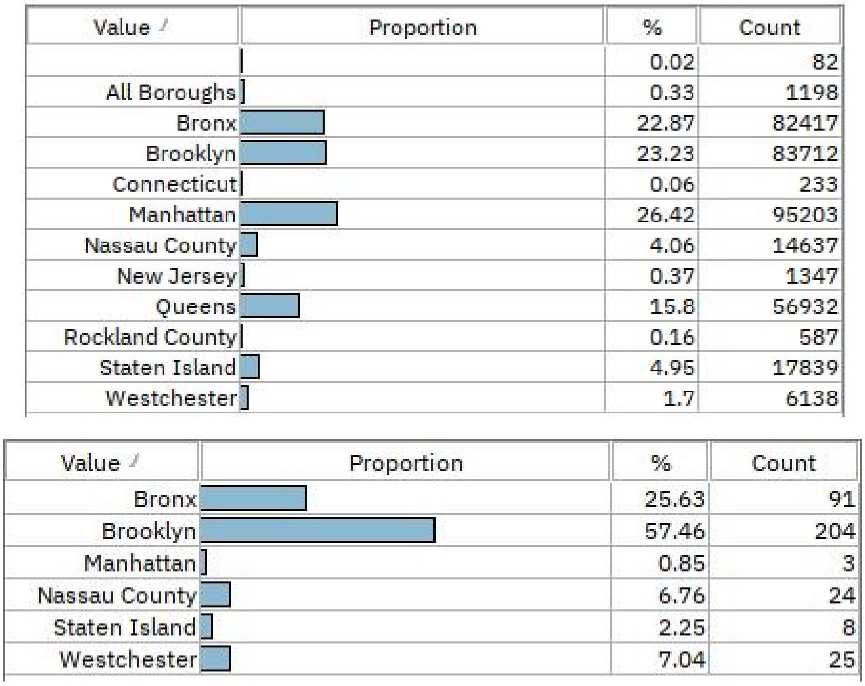


Figure 2b. Data Exploration of Factors' Effect on Delay Time



| Value | Proportion | % | Count |
|---|---|---|---|
| | | 0.02 | 82 |
| All Boroughs | | 0.33 | 1198 |
| Bronx | | 22.87 | 82417 |
| Brooklyn | | 23.23 | 83712 |
| Connecticut | | 0.06 | 233 |
| Manhattan | | 26.42 | 95203 |
| Nassau County | | 4.06 | 14637 |
| New Jersey | | 0.37 | 1347 |
| Queens | | 15.8 | 56932 |
| Rockland County | | 0.16 | 587 |
| Staten Island | | 4.95 | 17839 |
| Westchester | | 1.7 | 6138 |

| Value | Proportion | % | Count |
|---|---|---|---|
| Bronx | | 25.63 | 91 |
| Brooklyn | | 57.46 | 204 |
| Manhattan | | 0.85 | 3 |
| Nassau County | | 6.76 | 24 |
| Staten Island | | 2.25 | 8 |
| Westchester | | 7.04 | 25 |

Figure 2c. Data Exploration of Difference via Incident Issues

In the data preprocessing step, we clean the data by assigning null delay time for a consistent data including both breakdowns and delays, and removing other Nulls for reducing the interaction noise. We also deal with the outliers based on the standard deviation. We do need to consider the issue of using the test phase data as we have not applied the mean or mode replacement. The outlier treatment is not religious, but the result has not changed because the outlier due to manual error is extremely higher than the right records. We format the timestamp data and derive some new features such as hour, day of week, month, report response time for further data analysis. We integrate the dataset of different years. And we select the data features based on our hypothesis after data exploration as we recognize that not all features are equally relevant to our analysis like ID data fields not contributing significantly to our prediction goals or timestamp data fields from which we have derived new features. We can focus on the remaining data fields. For the first hypothesis, our derived frequencies of different categorical features are our main selections, and for the second hypothesis, the dummy vectors of different categorical features are our main selections. In the data transformation step, we decide our target value based on different presuppositions, and use the feature selection model in IBM SPSS Modeler or apply basic regression models in Python for data reduction. We apply a logarithmic transformation in our numeric data for getting a normal-like distribution.

After these steps, we choose regression for both purposes. First, we have a supervised learning partition. The inherent ability of regression to predict continuous numerical outcomes or aim to unravel the intricate relationships between our features and the frequency is a perfect match. Also, in our quest to predict delay times and uncover the reasons behind service delays, we settle on using regression analysis for our task.

## 4.2. Implementation Technology

We first conduct exploratory analysis and discussion using different first data mining algorithms. Our attempted algorithms are as following: a) linear regression including basic linear regression, generalised linear regression, Lasso regression, mulitple regression, b) decision tree including basic decision tree, C&R tree, c)ensemble tree methods including random forest, d) support vector machine, and e) basic neural network. At the first exploratory analysis, we find different models provide different insights, like the importance ranking of factors change. So, we decide which model to choose mainly based on root mean square error, p-value of different factors and efficiency of dealing with the large dataset for a acceptable result, along with the interpretability.

For the first hypothesis, we choose C&R regression tree model, which recursively divides the dataset into subsets in a way that minimizes the variance of our target frequency. We choose different hyperparameters to prune the from overfitting and also try boosting methods for the most an appropriate model. For the second hypothesis, we choose generalised linear regression, which models the relationship between a dependent response following different probability distributions and some independent predictors using a linear equation. We compare the prediction accuracy scores in different link functions and probability distributions, and use k-fold cross validation to tune the hyperparameters for getting the final model.

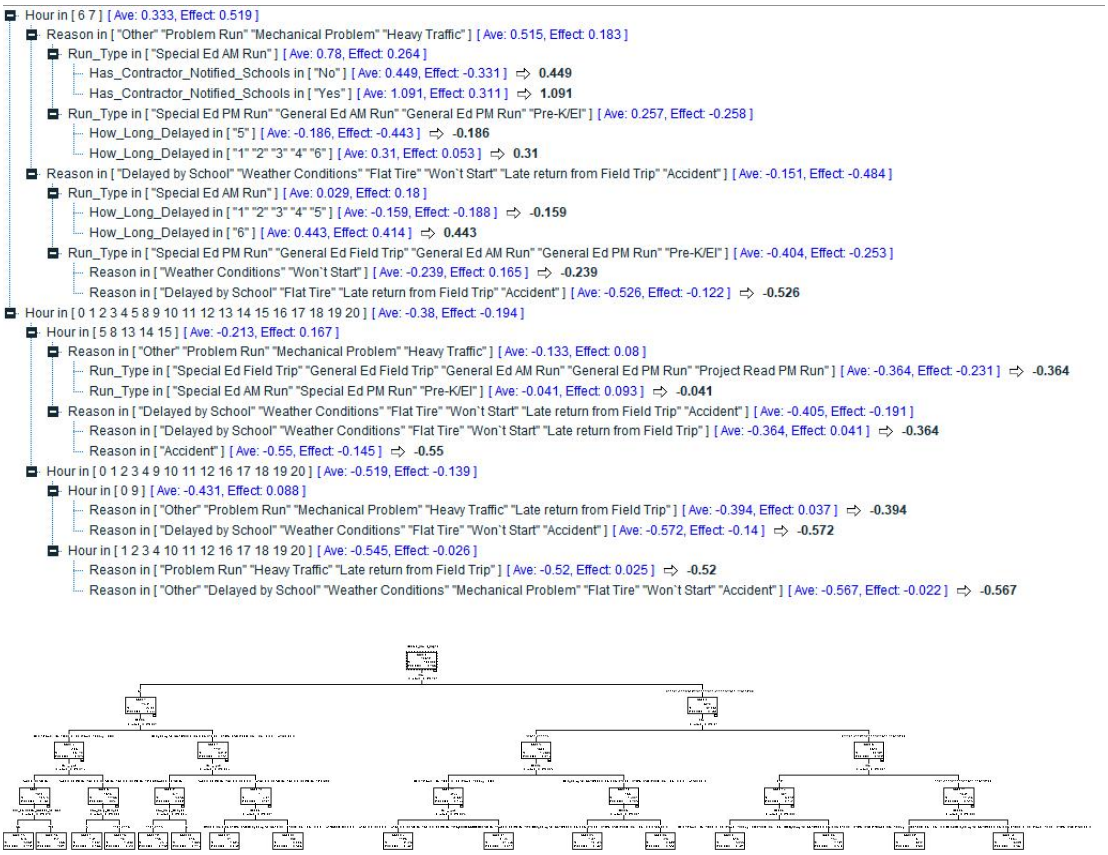
We generate the training set and testing set. Here we set the training partition size as 85% and the testing partition size as 15%. As we have chosen the model and the parameters, we do not need a validation set. The partition uses a raondom seed. Our training phase will be used to train the model, and the testing phase will be used to test the model performance. After conducting the data mining process, we get the output, visualise and interpret the results of patterns for valuable insights.

# 5. Result

## 5.1. Factor Effect on Delay/Breakdown Frequency Based on C&R Tree

The detailed tree structure is shown in Fig 3. First, there exists a special pattern in hour 6 and 7, which is reasonable because it is when students take a bus to the school. As for the morning rush data, the model uses the second important feature reason to make a classify. Here, heavy traffic, mechanical problem and problem run are one group. We can regard this group as vehicle problems. Keep going to the leaf node. Most data records belong to the special ED AM Run. As it is the most common method for bus vendors, there is no child of it. And for the other types of run type, the breakdown rate is similar with a slight delay rather than a long delay duration. As for the other reason node, the model still tracks the run type first. And then it is classfied with the reason again. Here clearly the two nodes are divided into two groups, one group having the reason naturally, and one group having the reason manually.

Let’s turn to the other node after the first hour division. This side extract spcial hour period first. From it we know the rush hours’ ranking. Again, the second par of each is classified by reason. And two reason groups can almost be regarded as heavy traffic problems along with other vehicle problems and the other. Here we can still find some difference from the previous nodes. The run types are divided by ‘curb-to-curb’, too, but they do not care the morning or afternoon and picked children. As for the delay type, the breakdown type shows a clear difference compared to the others.



```
⊟ Hour in [ 6 7 ] [ Ave: 0.333, Effect: 0.519 ]
  ⊟ Reason in [ "Other" "Problem Run" "Mechanical Problem" "Heavy Traffic" ] [ Ave: 0.515, Effect: 0.183 ]
    ⊟ Run_Type in [ "Special Ed AM Run" ] [ Ave: 0.78, Effect: 0.264 ]
        Has_Contractor_Notified_Schools in [ "No" ] [ Ave: 0.449, Effect: -0.331 ] ⇨ 0.449
        Has_Contractor_Notified_Schools in [ "Yes" ] [ Ave: 1.091, Effect: 0.311 ] ⇨ 1.091
    ⊟ Run_Type in [ "Special Ed PM Run" "General Ed AM Run" "General Ed PM Run" "Pre-K/EI" ] [ Ave: 0.257, Effect: -0.258 ]
        How_Long_Delayed in [ "5" ] [ Ave: -0.186, Effect: -0.443 ] ⇨ -0.186
        How_Long_Delayed in [ "1" "2" "3" "4" "6" ] [ Ave: 0.31, Effect: 0.053 ] ⇨ 0.31
  ⊟ Reason in [ "Delayed by School" "Weather Conditions" "Flat Tire" "Won`t Start" "Late return from Field Trip" "Accident" ] [ Ave: -0.151, Effect: -0.484 ]
    ⊟ Run_Type in [ "Special Ed AM Run" ] [ Ave: 0.029, Effect: 0.18 ]
        How_Long_Delayed in [ "1" "2" "3" "4" "5" ] [ Ave: -0.159, Effect: -0.188 ] ⇨ -0.159
        How_Long_Delayed in [ "6" ] [ Ave: 0.443, Effect: 0.414 ] ⇨ 0.443
    ⊟ Run_Type in [ "Special Ed PM Run" "General Ed Field Trip" "General Ed AM Run" "General Ed PM Run" "Pre-K/EI" ] [ Ave: -0.404, Effect: -0.253 ]
        Reason in [ "Weather Conditions" "Won`t Start" ] [ Ave: -0.239, Effect: 0.165 ] ⇨ -0.239
        Reason in [ "Delayed by School" "Flat Tire" "Late return from Field Trip" "Accident" ] [ Ave: -0.526, Effect: -0.122 ] ⇨ -0.526
⊟ Hour in [ 0 1 2 3 4 5 8 9 10 11 12 13 14 15 16 17 18 19 20 ] [ Ave: -0.38, Effect: -0.194 ]
  ⊟ Hour in [ 5 8 13 14 15 ] [ Ave: -0.213, Effect: 0.167 ]
    ⊟ Reason in [ "Other" "Problem Run" "Mechanical Problem" "Heavy Traffic" ] [ Ave: -0.133, Effect: 0.08 ]
        Run_Type in [ "Special Ed Field Trip" "General Ed Field Trip" "General Ed AM Run" "General Ed PM Run" "Project Read PM Run" ] [ Ave: -0.364, Effect: -0.231 ] ⇨ -0.364
        Run_Type in [ "Special Ed AM Run" "Special Ed PM Run" "Pre-K/EI" ] [ Ave: -0.041, Effect: 0.093 ] ⇨ -0.041
    ⊟ Reason in [ "Delayed by School" "Weather Conditions" "Flat Tire" "Won`t Start" "Late return from Field Trip" "Accident" ] [ Ave: -0.405, Effect: -0.191 ]
        Reason in [ "Delayed by School" "Weather Conditions" "Flat Tire" "Won`t Start" "Late return from Field Trip" ] [ Ave: -0.364, Effect: 0.041 ] ⇨ -0.364
        Reason in [ "Accident" ] [ Ave: -0.55, Effect: -0.145 ] ⇨ -0.55
  ⊟ Hour in [ 0 1 2 3 4 9 10 11 12 16 17 18 19 20 ] [ Ave: -0.519, Effect: -0.139 ]
    ⊟ Hour in [ 0 9 ] [ Ave: -0.431, Effect: 0.088 ]
        Reason in [ "Other" "Problem Run" "Mechanical Problem" "Heavy Traffic" "Late return from Field Trip" ] [ Ave: -0.394, Effect: 0.037 ] ⇨ -0.394
        Reason in [ "Delayed by School" "Weather Conditions" "Flat Tire" "Won`t Start" "Accident" ] [ Ave: -0.572, Effect: -0.14 ] ⇨ -0.572
    ⊟ Hour in [ 1 2 3 4 10 11 12 16 17 18 19 20 ] [ Ave: -0.545, Effect: -0.026 ]
        Reason in [ "Problem Run" "Heavy Traffic" "Late return from Field Trip" ] [ Ave: -0.52, Effect: 0.025 ] ⇨ -0.52
        Reason in [ "Other" "Delayed by School" "Weather Conditions" "Mechanical Problem" "Flat Tire" "Won`t Start" "Accident" ] [ Ave: -0.567, Effect: -0.022 ] ⇨ -0.567
```

Figure 3. C&R Tree Structure

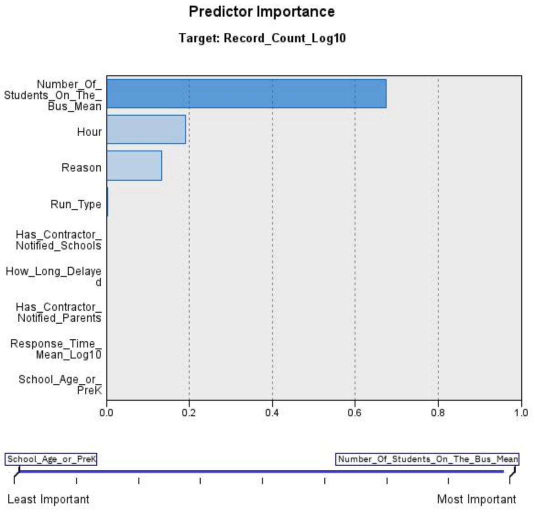


Figure 4. Factor Importance Ranking of C&R Tree

Our mined patterns are highly related to the four main predictors (see in Fig 4), data records, at which time in a day, with which reason, run type and delay duration, can be regarded as having similar delay or breakdown induced incidents' occurence times (see interactions in Fig 5, see effect on frequency in Fig 6). We can learn from that there are some rush hours. Different delay/breakdown reasons tends to occur at different hour. For example, there is a high rate at 0 for problem run, which might be the reason that the bus has some problem and the bus vendor reports it at that time. For another example, the late return from field trip tends to occur at afternoon, which is reasonable because it should be the bus's second trip that day.

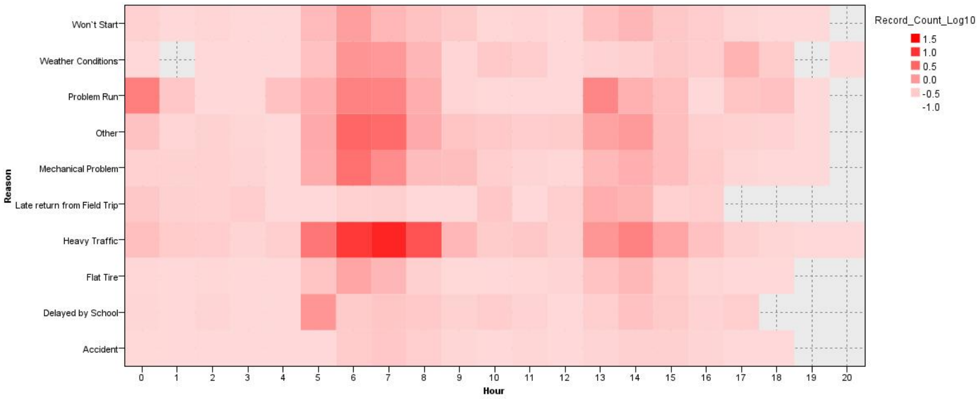


Figure 5a. Hour and Reason Distribution

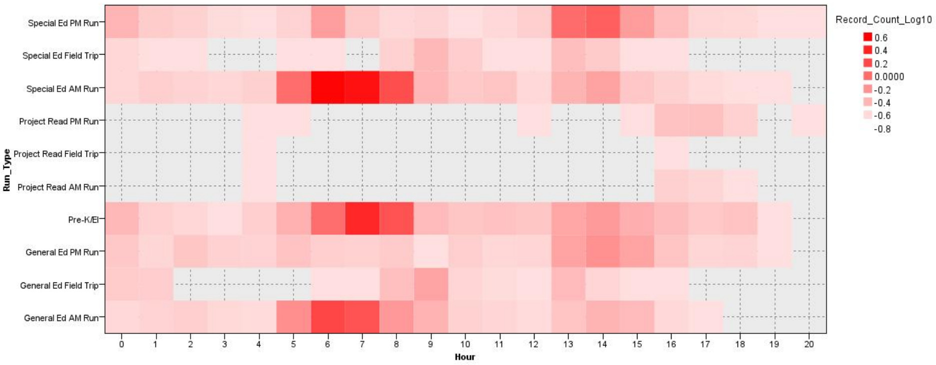


Figure 5b. Hour and Run Type Distribution

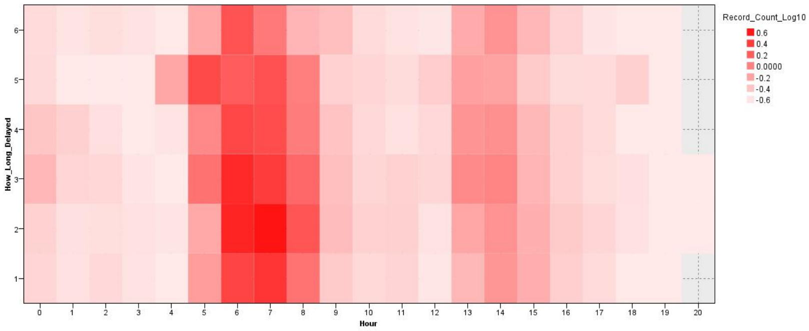


Figure 5c. Hour and Delay Duration Distribution

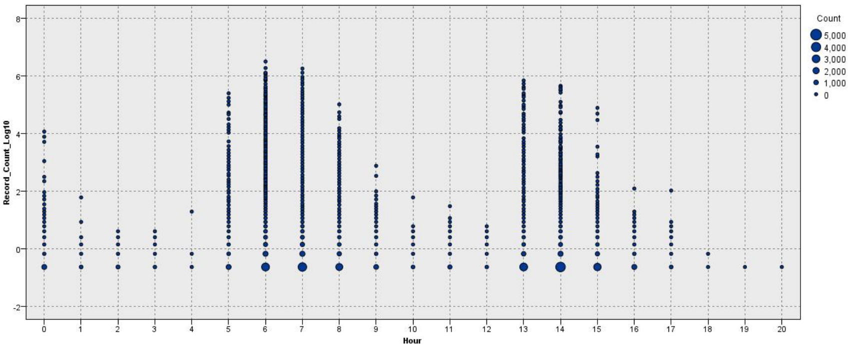


Figure 6a. Hour Effect on Frequency

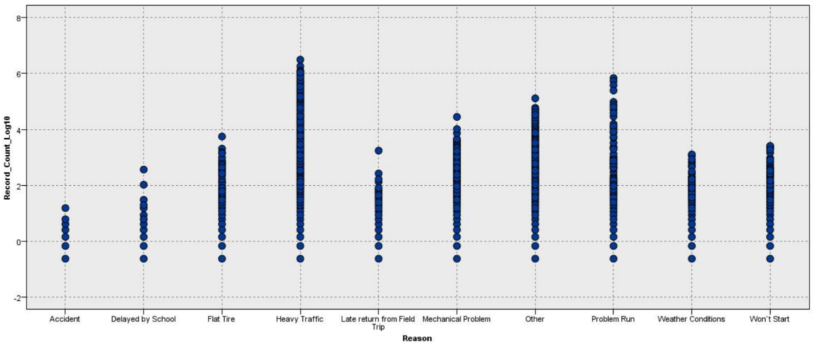


Figure 6b. Reason Effect on Frequency

During our comprehensive pattern analysis, several intriguing insights emerged, shedding light on the underlying dynamics of incidents related to bus delays and breakdowns. A significant finding pertains to the distinctive pattern observed during hours 6 and 7. This alignment with the early morning school commute resonates logically, as these hours coincide with when students typically embark on their bus journeys to school.

As we delved into the morning rush data, a noteworthy pattern emerged. The model strategically leveraged the 'reason' parameter as a key classifier. Within this parameter, a convergence of factors such as heavy traffic, mechanical issues, and problematic runs formed a discernible cluster, aptly labeled as 'vehicle problems'. This classification underscores a crucial link between these reasons and their contribution to vehicular challenges.

Further exploring the pattern, a particular leaf node stood out—dominated by instances of the 'Special ED AM Run'. This finding aligns with the fact that this method is the most frequently employed by bus vendors, thus explaining the absence of a child node. Surprisingly, other run types exhibited similar breakdown rates, albeit with minor delays instead of prolonged durations.

The exploration extended to nodes connected to other reasons, unveiling an intricate layering of run types and reason classifications. This intricate division unveiled two distinct categories—one with reasons naturally aligned and another attributed manually.

A novel observation arose when investigating nodes following the initial hour-based division. These nodes spotlighted specific temporal periods of significance, effectively ranking rush hours. Notably, the second part of these periods was classified based on reason, revealing two distinct clusters—those associated with heavy traffic or vehicle problems and others belonging to a broader category.

This analysis also exposed differences from preceding nodes. The classification of run types in these segments expanded to include 'curb-to-curb', irrespective of morning or afternoon specifics or pick-up details. Furthermore, the categorization of delay types, particularly for breakdowns, exhibited marked differences compared to other factors.

In summary, our pattern analysis provides valuable insights into the intricate web of factors influencing bus delays and breakdowns. These patterns manifest at specific times, under particular reasons, run types, and delay durations. This understanding not only aids in predicting incidents but also informs proactive measures to address these challenges effectively.

We know that the outcome has a 2.8 mean error, around 30% accuracy score in one decimal digit rounding and over 90% accuracy score in digit rounding. However, according to the gain plot in Fig 7, we see the prediction increases the gain, which suggests that the model tends to over-predict the occurrence of delays, especially the overestimation of the problem's severity on the side of higher delay counts. The model consistently produces predicitons that are higher than the actual delay frequency. Such a pattern suggests a systematic bias in the model's predictions. These observations require us to include more relevant predictors, no matter whether in the dataset or not to capture the true patterns of incidents, and evaluate and address potential biases in the data. Other modeling algorithms should also be considered. Still, considering the complexity of the real world, we get meaningful insights.

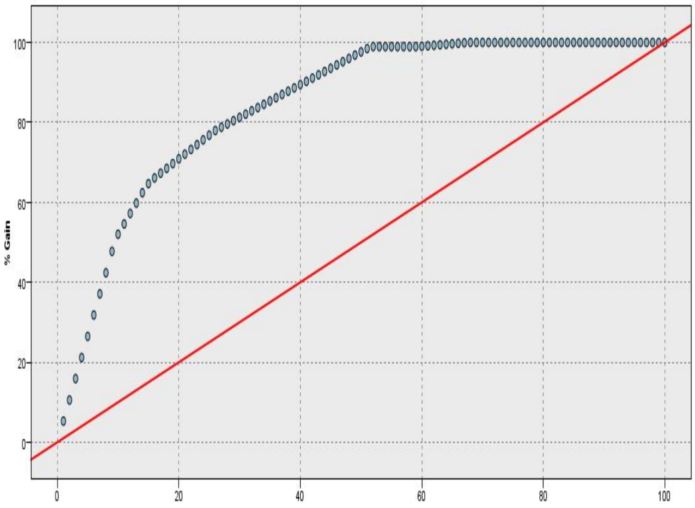
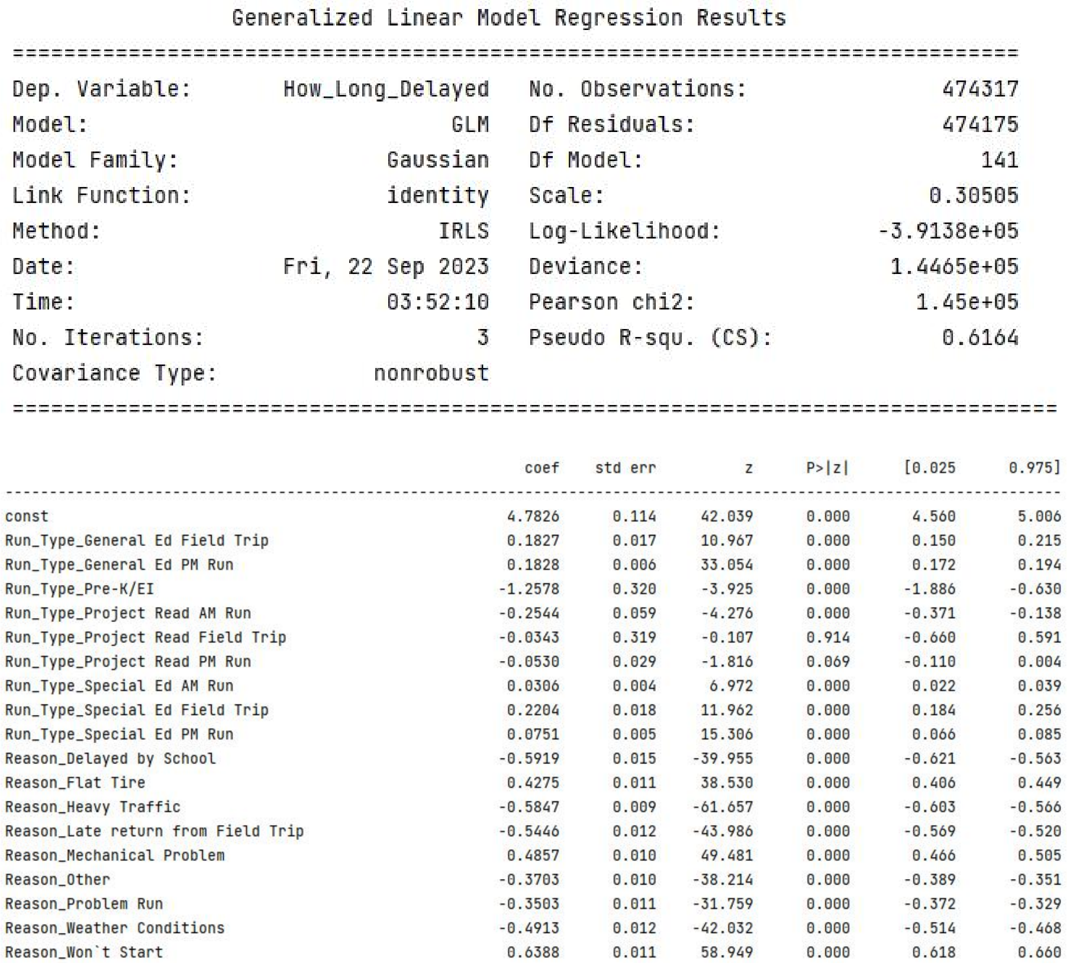


Figure 7. Gain Plot (right) of C&R Tree

### 5.2. Factor Effect on Duration of Delays Based on GLM

Our output of the model (see in Fig 8) for each coefficient has five values. The coefficient indicates the strength and direction of the relationship between the predictor variable and the response variable. The mean of distribution of $delay\ time \sim beta_0 + beta_1 * x_1 + \ldots\ beta_k * x_k$. Here, beta represents the coefficient we get in the summary, x indicates the factor. In our model, the factor is not the reason or other data fields themselves. The factor is the difference from a base reason level to a specific reason level. So we have more than 4 factors, the number of features we have selected.



```
                 Generalized Linear Model Regression Results
==============================================================================
Dep. Variable:       How_Long_Delayed   No. Observations:               474317
Model:                            GLM   Df Residuals:                   474175
Model Family:                Gaussian   Df Model:                          141
Link Function:               identity   Scale:                         0.30505
Method:                          IRLS   Log-Likelihood:            -3.9138e+05
Date:                Fri, 22 Sep 2023   Deviance:                   1.4465e+05
Time:                        03:52:10   Pearson chi2:                 1.45e+05
No. Iterations:                     3   Pseudo R-squ. (CS):             0.6164
Covariance Type:            nonrobust
=======================================================================================

                                                    coef    std err          z      P>|z|      [0.025      0.975]
--------------------------------------------------------------------------------------------------------------------
const                                             4.7826      0.114     42.039      0.000       4.560       5.006
Run_Type_General Ed Field Trip                    0.1827      0.017     10.967      0.000       0.150       0.215
Run_Type_General Ed PM Run                        0.1828      0.006     33.054      0.000       0.172       0.194
Run_Type_Pre-K/EI                                -1.2578      0.320     -3.925      0.000      -1.886      -0.630
Run_Type_Project Read AM Run                     -0.2544      0.059     -4.276      0.000      -0.371      -0.138
Run_Type_Project Read Field Trip                 -0.0343      0.319     -0.107      0.914      -0.660       0.591
Run_Type_Project Read PM Run                     -0.0530      0.029     -1.816      0.069      -0.110       0.004
Run_Type_Special Ed AM Run                        0.0306      0.004      6.972      0.000       0.022       0.039
Run_Type_Special Ed Field Trip                    0.2204      0.018     11.962      0.000       0.184       0.256
Run_Type_Special Ed PM Run                        0.0751      0.005     15.306      0.000       0.066       0.085
Reason_Delayed by School                         -0.5919      0.015    -39.955      0.000      -0.621      -0.563
Reason_Flat Tire                                  0.4275      0.011     38.530      0.000       0.406       0.449
Reason_Heavy Traffic                             -0.5847      0.009    -61.657      0.000      -0.603      -0.566
Reason_Late return from Field Trip               -0.5446      0.012    -43.986      0.000      -0.569      -0.520
Reason_Mechanical Problem                         0.4857      0.010     49.481      0.000       0.466       0.505
Reason_Other                                     -0.3703      0.010    -38.214      0.000      -0.389      -0.351
Reason_Problem Run                               -0.3503      0.011    -31.759      0.000      -0.372      -0.329
Reason_Weather Conditions                        -0.4913      0.012    -42.032      0.000      -0.514      -0.468
Reason_Won`t Start                                0.6388      0.011     58.949      0.000       0.618       0.660
```

Figure 8. GLM Summary (Boro and Bus Company Factors in Appendix)

As for each distinct value of the 'Run_Type' feature, the model has assigned a coefficient that indicates the weight of that feature in predicting the 'How_Long_Delayed' target variable. 'General Ed AM Run' has a coefficient of approximately 0.0154, indicating a slight positive effect on delays. 'General Ed Field Trip' has a coefficient of approximately -0.0582, suggesting a negative effect on delays. Similarly, the coefficients for the remaining in ‘Run_Type’ and for 'Reason', 'Boro', and 'Bus_Company_Name' provide insights into how these categories affect delays. As for the reason, the mechinical issues, tyre issues and ‘won’t start’ are the three positive reasons, which lead to a long delay time, and this model conforms with the result of the model of decision tree in our previous step. And the other reasons except the three lead to a short delay time. And obviously they are incidents we generally consider key to delay in our daily life. Not just within one data feature or one data field, by comparing those coefficients, we clearly find that the average of reason and bus type have a higher impact on delay time. As for boro or bus company name, we can find some extreme high or low (negative) values. So a marcro control for bus service by coordinating transportation and improving bus qualities and functions are important. At the same time, encourging good areas or companies and punishing the worse ones are also important. We can refer to our model coefficients to find those pinned areas or companies. We notice that our feature is a vector and each feature include an index and a encoder. If we just have the encoder, we can not prove any numeric relation. So with this format, we avoid a wrong sorting, which may bring noisy relationships. However, as for p values, our results are not as expected, and worse than we have done in the previous iteration. We see the strong effect with larger coefficient in run type and reason group are more convincable. Most coefficients for the boro are also be trusted. And if the p value is small for the company, the effect is high. Here we find that the impressive features are trustable, which may due to the data distribution that different types change the distribution of delay time obviously as we have more data entries to support the pattern.

And beta0 can be the intercept which means all the four data fields contain the data values of the basic level, and in this case, what the distribution of log(delay time) will be. In this linear regression model, the intercept represents the predicted value of the dependent variable 'How_Long_Delayed' when all predictor variables (features) are set to zero. So it refers to the model's inherent tendency to predict a certain value when no features are present. It provides an adjustment to account for the mean (around 3.7) or baseline value of the dependent variable. So our intercept is a bit lower, which indicates an inclination of less delay time. The intercept helps the model capture any systematic or inherent shifts in the target variable's values, separate from the influence of the predictors. For example, if we want to add some weight to some certain features, we can use the intercept as an offset to generate our survey data.

The RMSE of approximately 0.347 implies that, on average, our model's predictions have an error of 0.347 units when estimating how long a bus is delayed. So, if our model predicts a delay of 4.5 units (90 minutes), the actual delay is likely to be within the range of 4.2 units (65 minutes) to 4.8 units (125 minutes) on average. If the prediction is small, the effect of the error is small due to a log transformation. So it seems to perform better for grouping the delay time to our original data type. An R-squared value of approximately 0.465 means that our model accounts for about 46.52% of the variance in the delays. In other words, our model explains a moderate portion of the variation in bus delays. This suggests that there are other factors or variables not included in your model that also influence the delays, or it does not follow the assumption of linear regression. Considering we have lost about 10% from the previous result of similar model for the whole database, our model is sensitive to the quantity of the data. Because we are actully using millions of catergorical combinations to make the numeric prediction, so we hope an ideal situation that each combination has sufficient data entries (4 million total for now).

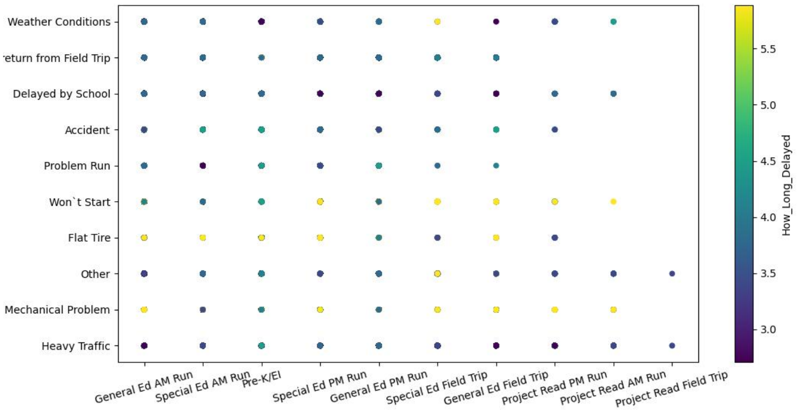


Figure 9a - Delay Time V.S. Reason and Run Type

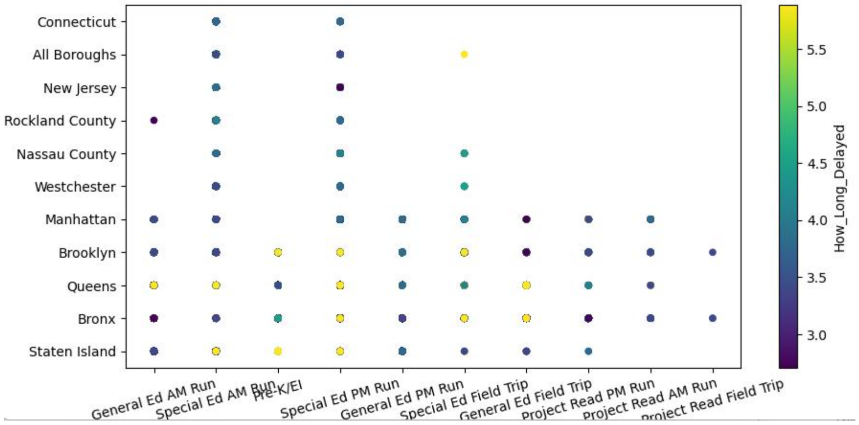


Figure 9b - Delay Time V.S. Boro and Run Type

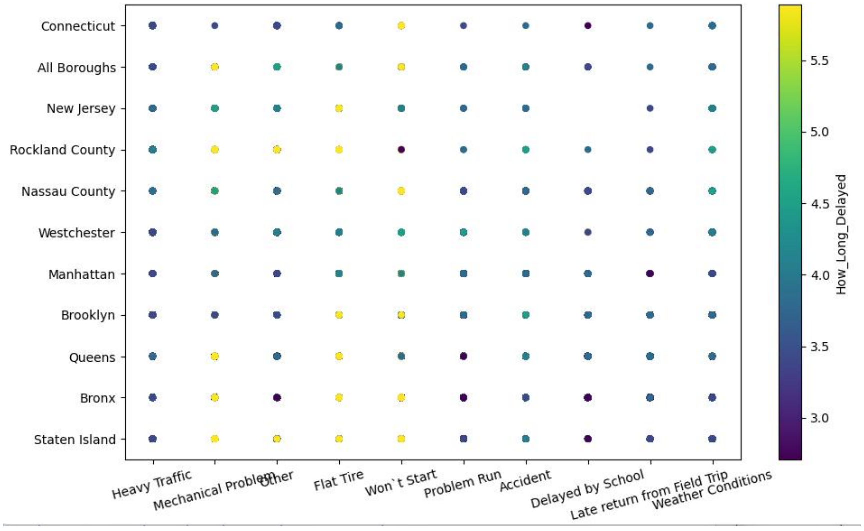


Figure 9c - Delay Time V.S. Reason and Boro

In this part, we will focus on the mined patterns (see in Fig 9). As mentioned, our mined patterns are highly related to the four main predictors, reason, run type, boro (geographical area), and bus company (we can not show in visualisation). They help us understand how these factors relate to delay times.

From the visualisations above, we can see the three predictors, reason, run type and boro do affect the delay time as we can see some combinations clearly lead to a high delay time with light yellow color and some lead to a low delay time with deep purple color.

Our model is to build a linear relationship between these four predictors and the delay time. So we can not only say the combination of these features leads to a high delay time, we can say how each feature affects the delay time. And if some features in different data fields all have positive effect on the delay time, the combination of these features leads to a high delay time. And if both two combinations leads to a high delay time, we can also compare which is higher through the coefficient estimator.

Let review the example we have mentioned above, a bus company like TRANS LLC, with a coefficient estimator of 1.47. This means that if you switch from the base bus company to TRANS LLC, you're 1.47 times more likely to experience a longer delay. It suggests that TRANS LLC may have service issues leading to longer delays. On the other hand, bus companies with negative coefficients tend to provide better service with shorter delays.We also look at different geographical areas (boros) and types of service runs. Positive coefficients suggest longer delays, while negative coefficients imply shorter delays. This information helps us identify which areas and service types are linked to longer or shorter delays, aiding in improving service quality.When it comes to reasons for delays, our model highlights three specific reasons—mechanical issues, tire problems, and 'won't start'—with positive coefficients, indicating longer delays. This aligns with our earlier decision tree model results. Other reasons, except for these three, lead to shorter delays.

We can get information about our model from figure 10. We know that our predicted values have a similar distribution with the that of the true values. At least our predicted values do not break the original distribution. But we can see a clear difference if the upper bound of the delay bound is over 90, because the largest class is what we manually set for the breakdown situation. So our model consider all the data entries no matter whether it is a delay type or breakdown type. Though it does not perform well from statistic angle, it is meaningful for our insights because a continuous prediction provides more information than the categorical prediction for our objective. In a normal Q-Q plot, the points should closely follow the diagonal line. Any deviations from this line may suggest departures from normality in the residuals. In our model, when the quantiles are over 2, the model performs worse due to the manually set data value.

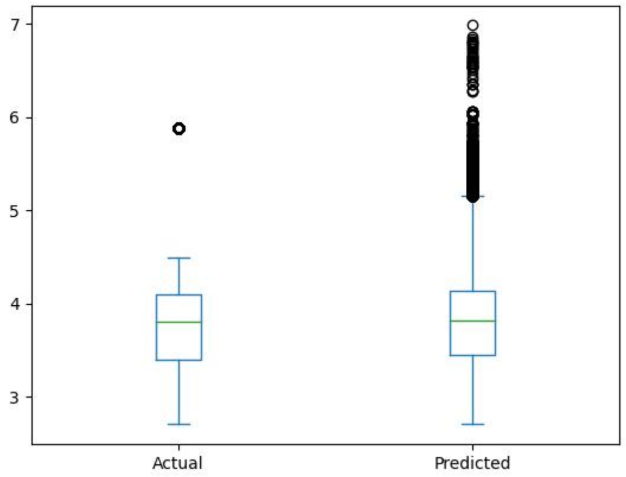


Figure 10a - Predictions vs Test Data

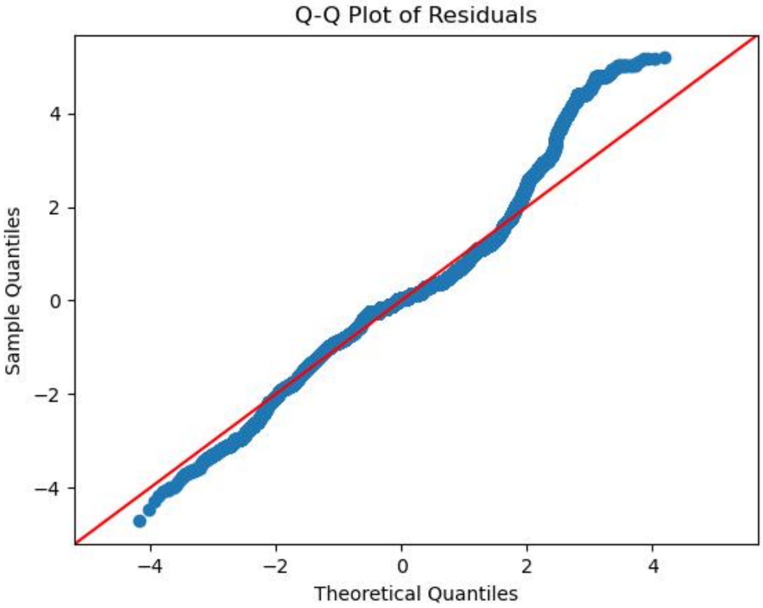


Figure 10b - Model Q-Q Plot

# 6. Discussion

Given the above insights, several proposed actions can be considered to improve the efficiency of school bus operations. First, we need an enhanced scheduling during peak hours. When students board the bus, transportation authorities can focus on optimizing schedules by adjusting the deployment of buses, optimizing routes, and ensuring timely pick-ups to accomodate the influx of students. Second, as there are many vehicle problem induced delays or breakdowns, it is important to have a early problem detection, real-time situation monitor, and regular maintenance schedules for mechanical issues. As for weather related issues or other reasons, the most effective way for improvement includes driver training, and continuous monitoring for data driven system suggestions. We can notice that the collaboration between transportation system, bus vendors and schools is vital, which should be responsible for the students’ experience on the school bus and parents’ worry for their children’s safety. The notification system relies on the manual work leading to another delay also needs to be improved. An open channel for feedback and communication might be a choice. We have found the bus company and boro factors effect. So a recommendation is that the authority needs to figure out some internal factors to bus companies with bad service performance and area managements with a higher rate delays that lead to potential, such as financial problem, technique scarcity or liability issues.

## 7. Future Work

Although the data analysis process provides the overall need for improvement in the New York school bus system, the model we get does not find some specific correlations or patterns, which means the importance of several factors or features have been addressed with strong confidence, while the causation pattern requires a deep investigation with more features.

Another important aspect is the safety related issues are absent because when we set if the delay or breakdown is caused by an incident or not (see in Fig 2c) as a predictor, we do not find a strong correlation between this feature and delays. However, delays caused by incidents are worth further research.

## 8. Conclusion

According to our goal of improving the efficiency and reliability of school bus services provided to both public and non-public schools in New York City by reducing the incidents caused by bus breakdowns and delays to ensure timely transportation for students, we use data mining techniques to extract insights from operational data, enabling informed decision-making and strategic planning. The rush hour correspponding to school commute, heavy traffic, mechanical vehicle problem, bus run type and internal factors of bus companies and areas contribute to either the frequency of delays/breakdowns or the duration of delays. These insights empower us to implement a data-driven approach, allowing for proactive measures in optimizing schedules during critical school commute hours. By addressing vehicle problems and specific run types with dedicated solutions, we aim to reduce the incidence of delays and enhance the reliability of school bus services. Additionally, considering internal factors of bus companies and areas, we can tailor our strategies to mitigate the impact of delays effectively, ensuring that students experience smoother and more dependable transportation services.

## Reference

[1] Selvam, M., Yadahalli, A. R., Dindi, M. M., & Nithin, B. (2022). Iot enabled school bus monitoring and notification system. In ICDSMLA 2020: Proceedings of the 2nd International

Conference on Data Science, Machine Learning and Applications (pp. 1205-1216). Springer Singapore.

[2] Mirchandani, P. B., Li, J. Q., & Hickman, M. (2010). A macroscopic model for integrating bus signal priority with vehicle rescheduling. Public Transport, 2, 159-172.

[3] Pejović, M. M., Denić, D. B., Pejović, M. M., Nešić, N. T., & Vasović, N. (2010). Microcontroller based system for electrical breakdown time delay measurement in gas-filled devices. Review of Scientific Instruments, 81(10).

[4] Borseti, E., & Berger, P. D. (2019). SCHOOL BUS BREAKDOWNS IN NEW YORK CITY.
Elizabeth Borseti, and Paul D. Berger. (2020). "SCHOOL BUS BREAKDOWNS IN NEW YORK CITY." International Journal of Research - Granthaalayah, 8(1), 336-349.

[5] Newton, R. M., & Thomas, W. H. (1969). Design of school bus routes by computer. Socio-Economic Planning Sciences, 3(1), 75-85.

[6] Burke, G. S., Lapidus, G. D., Zavoski, R. W., Wallace, L., & Banco, L. I. (1996). Evaluation of the effectiveness of a pavement stencil in promoting safe behavior among elementary school children boarding school buses. Pediatrics, 97(4), 520-523.

[7] Braca, J., Bramel, J., Posner, B., & Simchi-Levi, D. (1997). A computerized approach to the New York Cityschool bus routing problem. IIE transactions, 29, 693-702.

[8] Park, J., & Kim, B. I. (2010). The school bus routing problem: A review. European Journal of operational research, 202(2), 311-319.

[9] Shiyao, C., Yunyang, X., Baihui, L., & Hui, W. (2014, January). The Application of ZigBee Technology to the Intelligent Bus Query System. In 2014 Sixth International Conference on Measuring Technology and Mechatronics Automation (pp. 672-675). IEEE.

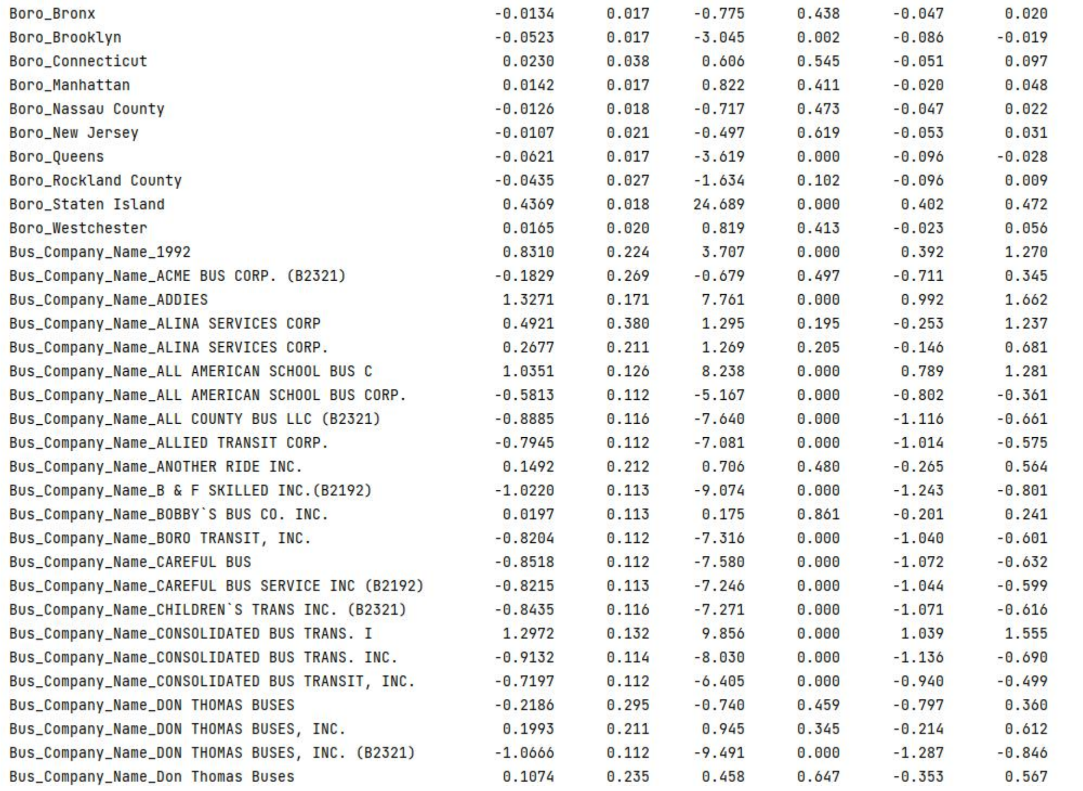
[10] Liu, Y., & Peng, H. (2019, April). Alarm system design of young children being left on school bus based on pressure sensor array. In IOP Conference Series: Materials Science and Engineering (Vol. 490, No. 7, p. 072061). IOP Publishing.

[11] Yue, W. S., Hoy, C. W., & Chye, K. K. (2017, October). A preliminary survey analysis of school shuttle bus system towards smart mobility solutions. In AIP Conference Proceedings (Vol. 1891, No. 1). AIP Publishing.

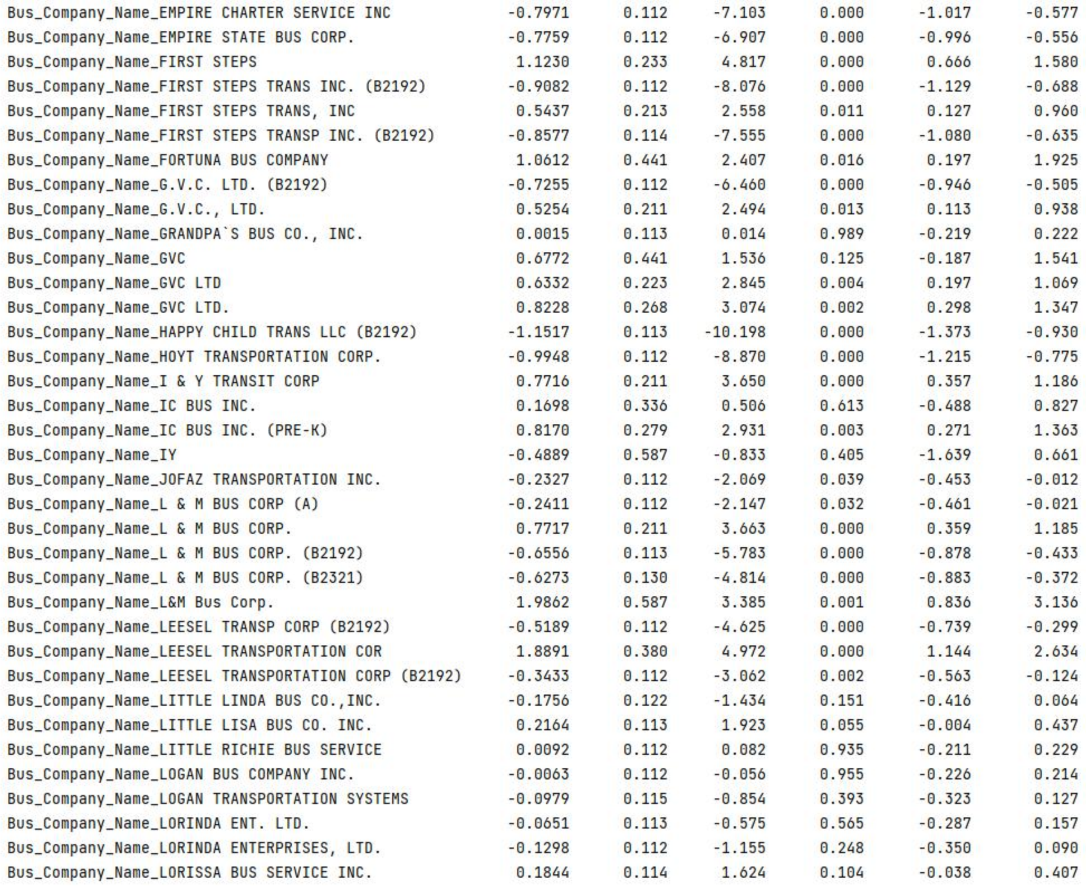
[12] Schröer, C., Kruse, F., & Gómez, J. M. (2021). A systematic literature review on applying CRISP-DM process model. Procedia Computer Science, 181, 526-534.

# Appendix

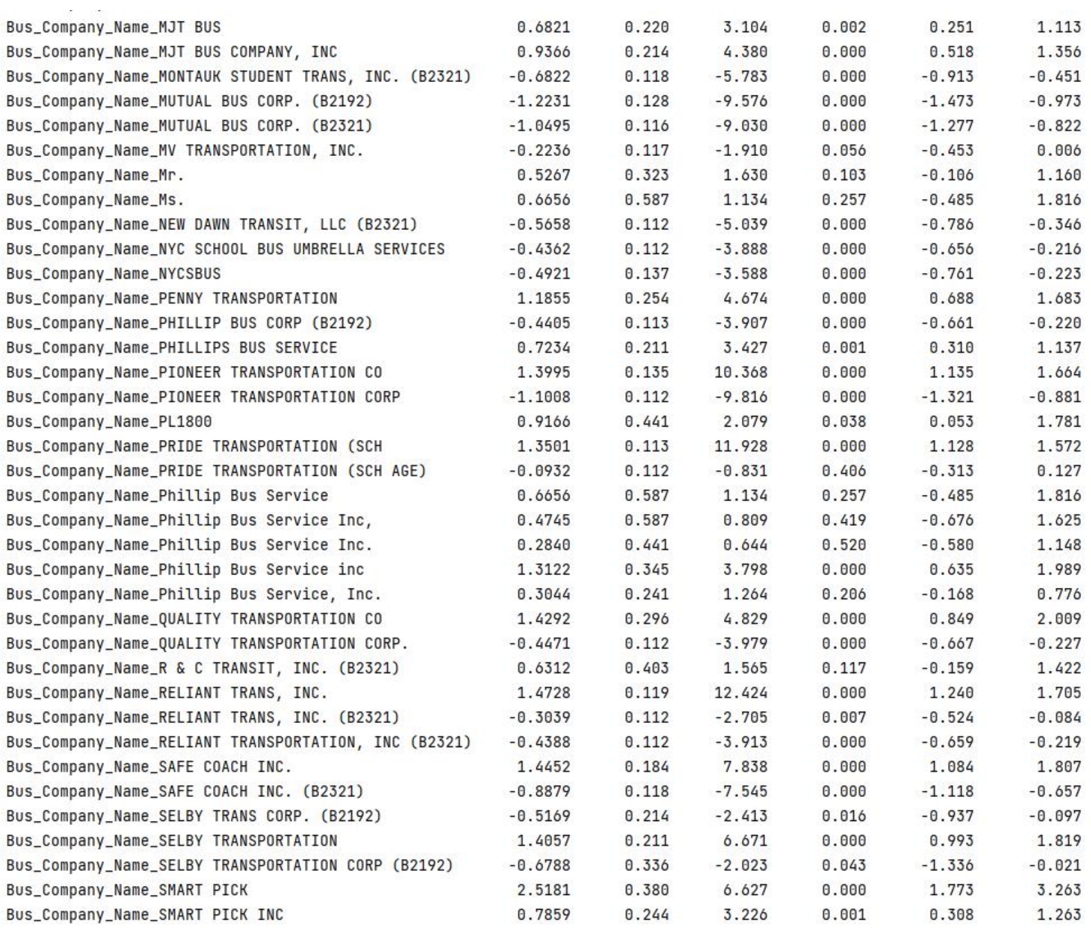
GLM Summary of Boro and Bus Company Factors



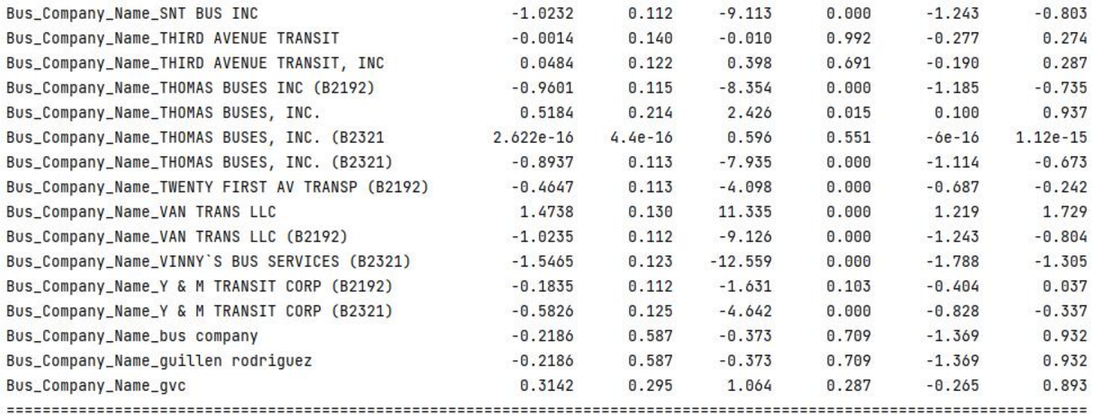
```
Boro_Bronx                                          -0.0134    0.017    -0.775    0.438    -0.047     0.020
Boro_Brooklyn                                       -0.0523    0.017    -3.045    0.002    -0.086    -0.019
Boro_Connecticut                                     0.0230    0.038     0.606    0.545    -0.051     0.097
Boro_Manhattan                                       0.0142    0.017     0.822    0.411    -0.020     0.048
Boro_Nassau County                                  -0.0126    0.018    -0.717    0.473    -0.047     0.022
Boro_New Jersey                                     -0.0107    0.021    -0.497    0.619    -0.053     0.031
Boro_Queens                                         -0.0621    0.017    -3.619    0.000    -0.096    -0.028
Boro_Rockland County                                -0.0435    0.027    -1.634    0.102    -0.096     0.009
Boro_Staten Island                                   0.4369    0.018    24.689    0.000     0.402     0.472
Boro_Westchester                                     0.0165    0.020     0.819    0.413    -0.023     0.056
Bus_Company_Name_1992                                0.8310    0.224     3.707    0.000     0.392     1.270
Bus_Company_Name_ACME BUS CORP. (B2321)             -0.1829    0.269    -0.679    0.497    -0.711     0.345
Bus_Company_Name_ADDIES                              1.3271    0.171     7.761    0.000     0.992     1.662
Bus_Company_Name_ALINA SERVICES CORP                 0.4921    0.380     1.295    0.195    -0.253     1.237
Bus_Company_Name_ALINA SERVICES CORP.                0.2677    0.211     1.269    0.205    -0.146     0.681
Bus_Company_Name_ALL AMERICAN SCHOOL BUS C           1.0351    0.126     8.238    0.000     0.789     1.281
Bus_Company_Name_ALL AMERICAN SCHOOL BUS CORP.      -0.5813    0.112    -5.167    0.000    -0.802    -0.361
Bus_Company_Name_ALL COUNTY BUS LLC (B2321)         -0.8885    0.116    -7.640    0.000    -1.116    -0.661
Bus_Company_Name_ALLIED TRANSIT CORP.               -0.7945    0.112    -7.081    0.000    -1.014    -0.575
Bus_Company_Name_ANOTHER RIDE INC.                   0.1492    0.212     0.706    0.480    -0.265     0.564
Bus_Company_Name_B & F SKILLED INC.(B2192)          -1.0220    0.113    -9.074    0.000    -1.243    -0.801
Bus_Company_Name_BOBBY`S BUS CO. INC.                0.0197    0.113     0.175    0.861    -0.201     0.241
Bus_Company_Name_BORO TRANSIT, INC.                 -0.8204    0.112    -7.316    0.000    -1.040    -0.601
Bus_Company_Name_CAREFUL BUS                        -0.8518    0.112    -7.580    0.000    -1.072    -0.632
Bus_Company_Name_CAREFUL BUS SERVICE INC (B2192)    -0.8215    0.113    -7.246    0.000    -1.044    -0.599
Bus_Company_Name_CHILDREN`S TRANS INC. (B2321)      -0.8435    0.116    -7.271    0.000    -1.071    -0.616
Bus_Company_Name_CONSOLIDATED BUS TRANS. I           1.2972    0.132     9.856    0.000     1.039     1.555
Bus_Company_Name_CONSOLIDATED BUS TRANS. INC.       -0.9132    0.114    -8.030    0.000    -1.136    -0.690
Bus_Company_Name_CONSOLIDATED BUS TRANSIT, INC.     -0.7197    0.112    -6.405    0.000    -0.940    -0.499
Bus_Company_Name_DON THOMAS BUSES                   -0.2186    0.295    -0.740    0.459    -0.797     0.360
Bus_Company_Name_DON THOMAS BUSES, INC.              0.1993    0.211     0.945    0.345    -0.214     0.612
Bus_Company_Name_DON THOMAS BUSES, INC. (B2321)     -1.0666    0.112    -9.491    0.000    -1.287    -0.846
Bus_Company_Name_Don Thomas Buses                    0.1074    0.235     0.458    0.647    -0.353     0.567
```



```
Bus_Company_Name_EMPIRE CHARTER SERVICE INC          -0.7971    0.112    -7.103    0.000    -1.017    -0.577
Bus_Company_Name_EMPIRE STATE BUS CORP.              -0.7759    0.112    -6.907    0.000    -0.996    -0.556
Bus_Company_Name_FIRST STEPS                          1.1230    0.233     4.817    0.000     0.666     1.580
Bus_Company_Name_FIRST STEPS TRANS INC. (B2192)      -0.9082    0.112    -8.076    0.000    -1.129    -0.688
Bus_Company_Name_FIRST STEPS TRANS, INC               0.5437    0.213     2.558    0.011     0.127     0.960
Bus_Company_Name_FIRST STEPS TRANSP INC. (B2192)     -0.8577    0.114    -7.555    0.000    -1.080    -0.635
Bus_Company_Name_FORTUNA BUS COMPANY                  1.0612    0.441     2.407    0.016     0.197     1.925
Bus_Company_Name_G.V.C. LTD. (B2192)                 -0.7255    0.112    -6.460    0.000    -0.946    -0.505
Bus_Company_Name_G.V.C., LTD.                         0.5254    0.211     2.494    0.013     0.113     0.938
Bus_Company_Name_GRANDPA`S BUS CO., INC.              0.0015    0.113     0.014    0.989    -0.219     0.222
Bus_Company_Name_GVC                                  0.6772    0.441     1.536    0.125    -0.187     1.541
Bus_Company_Name_GVC LTD                              0.6332    0.223     2.845    0.004     0.197     1.069
Bus_Company_Name_GVC LTD.                             0.8228    0.268     3.074    0.002     0.298     1.347
Bus_Company_Name_HAPPY CHILD TRANS LLC (B2192)       -1.1517    0.113   -10.198    0.000    -1.373    -0.930
Bus_Company_Name_HOYT TRANSPORTATION CORP.           -0.9948    0.112    -8.870    0.000    -1.215    -0.775
Bus_Company_Name_I & Y TRANSIT CORP                   0.7716    0.211     3.650    0.000     0.357     1.186
Bus_Company_Name_IC BUS INC.                          0.1698    0.336     0.506    0.613    -0.488     0.827
Bus_Company_Name_IC BUS INC. (PRE-K)                  0.8170    0.279     2.931    0.003     0.271     1.363
Bus_Company_Name_IY                                  -0.4889    0.587    -0.833    0.405    -1.639     0.661
Bus_Company_Name_JOFAZ TRANSPORTATION INC.           -0.2327    0.112    -2.069    0.039    -0.453    -0.012
Bus_Company_Name_L & M BUS CORP (A)                  -0.2411    0.112    -2.147    0.032    -0.461    -0.021
Bus_Company_Name_L & M BUS CORP.                      0.7717    0.211     3.663    0.000     0.359     1.185
Bus_Company_Name_L & M BUS CORP. (B2192)             -0.6556    0.113    -5.783    0.000    -0.878    -0.433
Bus_Company_Name_L & M BUS CORP. (B2321)             -0.6273    0.130    -4.814    0.000    -0.883    -0.372
Bus_Company_Name_L&M Bus Corp.                        1.9862    0.587     3.385    0.001     0.836     3.136
Bus_Company_Name_LEESEL TRANSP CORP (B2192)          -0.5189    0.112    -4.625    0.000    -0.739    -0.299
Bus_Company_Name_LEESEL TRANSPORTATION COR            1.8891    0.380     4.972    0.000     1.144     2.634
Bus_Company_Name_LEESEL TRANSPORTATION CORP (B2192)  -0.3433    0.112    -3.062    0.002    -0.563    -0.124
Bus_Company_Name_LITTLE LINDA BUS CO.,INC.           -0.1756    0.122    -1.434    0.151    -0.416     0.064
Bus_Company_Name_LITTLE LISA BUS CO. INC.             0.2164    0.113     1.923    0.055    -0.004     0.437
Bus_Company_Name_LITTLE RICHIE BUS SERVICE            0.0092    0.112     0.082    0.935    -0.211     0.229
Bus_Company_Name_LOGAN BUS COMPANY INC.              -0.0063    0.112    -0.056    0.955    -0.226     0.214
Bus_Company_Name_LOGAN TRANSPORTATION SYSTEMS        -0.0979    0.115    -0.854    0.393    -0.323     0.127
Bus_Company_Name_LORINDA ENT. LTD.                   -0.0651    0.113    -0.575    0.565    -0.287     0.157
Bus_Company_Name_LORINDA ENTERPRISES, LTD.           -0.1298    0.112    -1.155    0.248    -0.350     0.090
Bus_Company_Name_LORISSA BUS SERVICE INC.             0.1844    0.114     1.624    0.104    -0.038     0.407
```



```
Bus_Company_Name_MJT BUS                                 0.6821      0.220      3.104      0.002       0.251       1.113
Bus_Company_Name_MJT BUS COMPANY, INC                    0.9366      0.214      4.380      0.000       0.518       1.356
Bus_Company_Name_MONTAUK STUDENT TRANS, INC. (B2321)    -0.6822      0.118     -5.783      0.000      -0.913      -0.451
Bus_Company_Name_MUTUAL BUS CORP. (B2192)               -1.2231      0.128     -9.576      0.000      -1.473      -0.973
Bus_Company_Name_MUTUAL BUS CORP. (B2321)               -1.0495      0.116     -9.030      0.000      -1.277      -0.822
Bus_Company_Name_MV TRANSPORTATION, INC.                -0.2236      0.117     -1.910      0.056      -0.453       0.006
Bus_Company_Name_Mr.                                     0.5267      0.323      1.630      0.103      -0.106       1.160
Bus_Company_Name_Ms.                                     0.6656      0.587      1.134      0.257      -0.485       1.816
Bus_Company_Name_NEW DAWN TRANSIT, LLC (B2321)          -0.5658      0.112     -5.039      0.000      -0.786      -0.346
Bus_Company_Name_NYC SCHOOL BUS UMBRELLA SERVICES       -0.4362      0.112     -3.888      0.000      -0.656      -0.216
Bus_Company_Name_NYCSBUS                                -0.4921      0.137     -3.588      0.000      -0.761      -0.223
Bus_Company_Name_PENNY TRANSPORTATION                    1.1855      0.254      4.674      0.000       0.688       1.683
Bus_Company_Name_PHILLIP BUS CORP (B2192)               -0.4405      0.113     -3.907      0.000      -0.661      -0.220
Bus_Company_Name_PHILLIPS BUS SERVICE                    0.7234      0.211      3.427      0.001       0.310       1.137
Bus_Company_Name_PIONEER TRANSPORTATION CO               1.3995      0.135     10.368      0.000       1.135       1.664
Bus_Company_Name_PIONEER TRANSPORTATION CORP            -1.1008      0.112     -9.816      0.000      -1.321      -0.881
Bus_Company_Name_PL1800                                  0.9166      0.441      2.079      0.038       0.053       1.781
Bus_Company_Name_PRIDE TRANSPORTATION (SCH               1.3501      0.113     11.928      0.000       1.128       1.572
Bus_Company_Name_PRIDE TRANSPORTATION (SCH AGE)         -0.0932      0.112     -0.831      0.406      -0.313       0.127
Bus_Company_Name_Phillip Bus Service                     0.6656      0.587      1.134      0.257      -0.485       1.816
Bus_Company_Name_Phillip Bus Service Inc,                0.4745      0.587      0.809      0.419      -0.676       1.625
Bus_Company_Name_Phillip Bus Service Inc.                0.2840      0.441      0.644      0.520      -0.580       1.148
Bus_Company_Name_Phillip Bus Service inc                 1.3122      0.345      3.798      0.000       0.635       1.989
Bus_Company_Name_Phillip Bus Service, Inc.               0.3044      0.241      1.264      0.206      -0.168       0.776
Bus_Company_Name_QUALITY TRANSPORTATION CO               1.4292      0.296      4.829      0.000       0.849       2.009
Bus_Company_Name_QUALITY TRANSPORTATION CORP.           -0.4471      0.112     -3.979      0.000      -0.667      -0.227
Bus_Company_Name_R & C TRANSIT, INC. (B2321)             0.6312      0.403      1.565      0.117      -0.159       1.422
Bus_Company_Name_RELIANT TRANS, INC.                     1.4728      0.119     12.424      0.000       1.240       1.705
Bus_Company_Name_RELIANT TRANS, INC. (B2321)            -0.3039      0.112     -2.705      0.007      -0.524      -0.084
Bus_Company_Name_RELIANT TRANSPORTATION, INC (B2321)    -0.4388      0.112     -3.913      0.000      -0.659      -0.219
Bus_Company_Name_SAFE COACH INC.                         1.4452      0.184      7.838      0.000       1.084       1.807
Bus_Company_Name_SAFE COACH INC. (B2321)                -0.8879      0.118     -7.545      0.000      -1.118      -0.657
Bus_Company_Name_SELBY TRANS CORP. (B2192)              -0.5169      0.214     -2.413      0.016      -0.937      -0.097
Bus_Company_Name_SELBY TRANSPORTATION                    1.4057      0.211      6.671      0.000       0.993       1.819
Bus_Company_Name_SELBY TRANSPORTATION CORP (B2192)      -0.6788      0.336     -2.023      0.043      -1.336      -0.021
Bus_Company_Name_SMART PICK                              2.5181      0.380      6.627      0.000       1.773       3.263
Bus_Company_Name_SMART PICK INC                          0.7859      0.244      3.226      0.001       0.308       1.263
```



```
Bus_Company_Name_SNT BUS INC                            -1.0232      0.112     -9.113      0.000      -1.243      -0.803
Bus_Company_Name_THIRD AVENUE TRANSIT                   -0.0014      0.140     -0.010      0.992      -0.277       0.274
Bus_Company_Name_THIRD AVENUE TRANSIT, INC               0.0484      0.122      0.398      0.691      -0.190       0.287
Bus_Company_Name_THOMAS BUSES INC (B2192)               -0.9601      0.115     -8.354      0.000      -1.185      -0.735
Bus_Company_Name_THOMAS BUSES, INC.                      0.5184      0.214      2.426      0.015       0.100       0.937
Bus_Company_Name_THOMAS BUSES, INC. (B2321            2.622e-16    4.4e-16      0.596      0.551      -6e-16    1.12e-15
Bus_Company_Name_THOMAS BUSES, INC. (B2321)             -0.8937      0.113     -7.935      0.000      -1.114      -0.673
Bus_Company_Name_TWENTY FIRST AV TRANSP (B2192)         -0.4647      0.113     -4.098      0.000      -0.687      -0.242
Bus_Company_Name_VAN TRANS LLC                           1.4738      0.130     11.335      0.000       1.219       1.729
Bus_Company_Name_VAN TRANS LLC (B2192)                  -1.0235      0.112     -9.126      0.000      -1.243      -0.804
Bus_Company_Name_VINNY`S BUS SERVICES (B2321)           -1.5465      0.123    -12.559      0.000      -1.788      -1.305
Bus_Company_Name_Y & M TRANSIT CORP (B2192)             -0.1835      0.112     -1.631      0.103      -0.404       0.037
Bus_Company_Name_Y & M TRANSIT CORP (B2321)             -0.5826      0.125     -4.642      0.000      -0.828      -0.337
Bus_Company_Name_bus company                            -0.2186      0.587     -0.373      0.709      -1.369       0.932
Bus_Company_Name_guillen rodriguez                      -0.2186      0.587     -0.373      0.709      -1.369       0.932
Bus_Company_Name_gvc                                     0.3142      0.295      1.064      0.287      -0.265       0.893
========================================================================================================================
```